Supplemental Online Material: A Pollen "Forward" Model to Enhance the Realism of the BHM

S.1 Specification, testing, and estimation of the binomial-logistic GLM

In order to specify the particular pollen types to be used in the GLM ratio model described in the article text, a systematic examination of one-half of the data set (covering central to northeastern temperate North America, see Fig. S1) for candidate ratio models was done. The data come from the Modern Pollen Data Base for North American and Greenland (Whitmore et al. 2005). First, the 64 pollen types identified by Williams and Shuman (2008) as best for climate reconstruction in the region using the modern analog technique (MAT)* were screened for their correlations with July mean

* The MAT is a state-of-the-art method of using large-dimension multivariate (i.e., multitaxonomic) pollen data in environmental reconstruction. In a common form of this method, multivariate distances between pollen assemblages at modern data sites are compared against known vegetation characteristics associated with each assemblage, and deterministic or "fuzzy" distance thresholds are developed to distinguish pollen assemblages coming from similar vegetation (smaller distances) from those coming from dissimilar vegetation (larger distances). Test characterization techniques derived from other fields, such as receiver operating characteristic (ROC) analysis (Gavin et al. 2003; Wahl 2004), are sometimes used to optimize the discrimination of the identified distance threshold(s). These optimizations can be tuned to focus on reducing false positive or false negative errors (or to jointly minimize both errors) as considered appropriate for the analysis at hand. The analog thresholds identified are then applied to comparisons between the modern pollen assemblages and fossil pollen assemblages that have been chosen for use in environmental reconstruction. Positive modern-fossil analog selections from these comparisons are interpreted to mean that the vegetation and environmental conditions, including climate, at the time the fossil sample was deposited were the same as those at its modern analog site(s). If multiple modern

temperature (in terms of the monthly mean climatology over 1961-1990). The subsets of pollen types with stronger positive and negative correlations were then examined using a power set analysis; inputting all possible forms of pollen ratios definable from the subsets into the binomial logistic GLM to identify the better performing taxonomic models in terms of explained deviance and other analytics such as the AIC measure. Because ratio models with as few as two numerator taxa and two denominator-specific taxa showed performance nearly as good as that obtained by the "best" performing models identified mechanistically from the power set, a rule was applied to limit the selected ratio model to no more than these numbers of numerator and denominator-specific taxa. Within this partitioning, a further criterion of ecological relevance (including spatial distribution of the plants in the ratio) was applied, which led to two "semi-finalist" ratio models: spruce (*Picea*) plus birch (*Betula*) versus *Quercus* plus hickory (*Carya*); and *Picea* plus *Betula* versus *Quercus* plus elm (*Ulmus*). Due to the potential for bias in the calibration data due to the widespread prevalence of Dutch elm disease in the region during the 20th century, the former model was selected for final use, taking the form (*Picea*+*Betula*)/(*Picea*+*Betula*+*Quercus*+*Carya*). Ecologically, it is well-known that *Quercus* and *Carya* are characteristic trees of areas with relatively warm summers in this

---

analogs are selected, then the average of the modern sites' climate values is used as an approximate EV reconstruction of past climate conditions, also allowing an estimate of the uncertainty associated with this reconstruction. Weighting is often used in calculating these averages (and in their corresponding standard error calculations), frequently employing some kind of inverse distance scheme so that the "best" analogs carry the most weight in the reconstruction. Cf. Jackson and Williams (2004) for a general discussion of the MAT.

region and that *Picea* and *Betula* are characteristic trees of areas with relatively cool summers.

The *Picea+Betula* versus *Quercus +Carya* model was then re-estimated using the half of the data not employed in the power set/ecological selection process. The results are shown in Figure S1 (note that the fit on this second half of the data is highly similar to that obtained on the first half of the data in the power set analysis). It should be noted that sites with values of 1 or 0 for the pollen ratio (1.0%) were excluded from the GLM estimation under the assumption that these extreme cases represent "saturation" of the ratio model; i.e., the vegetation at the site does not have sensitivity to variations in temperature that can be represented by the model. Similarly, only fossil sites whose ratios are bounded by 0.9 and 0.1 have been examined for paleoreconstruction (see section S.2 below), so that near-saturated portions of the fitted models are avoided as likely carrying relatively weak climatic information. Sites with July mean temperature less than 16° (4.4%) were also excluded, as representing regionally anomalous cold sites with similarly anomalous vegetation. Sites whose primary data are reported as percent or per mil values (16.6%) were not utilized because the binomial logistic GLM requires count data. The fitted *Picea+Betula* versus *Quercus +Carya* model has explained deviance (denoted as R^2, analogous to explained variation in a standard linear model) of 0.792.

The R^2 value of 0.792 represents very high explained deviance/explained variation for a statistical ecological model; the highest value obtained in the power set analysis for this metric was ~0.81. Employing the full 64-taxa Williams-Shuman data set in an approximately comparable MAT validation test (reconstructing each site's

temperature using analogs selected from all the other sites) for a slightly smaller region (41-50° N, 60-92° W) yielded an explained variation of 0.741. Thus, taxonomic reduction by a factor of 16 in the binomial-logistic ratio model performs as well or better than a taxonomically optimized, state-of-the-art MAT. This is an unexpected result, at least in terms of magnitude of the reduction factor, and suggests that the binomial-logistic forward ratio model has greater capacity than the MAT to distinguish the temperature signal carried by a few dominant pollen types, at least in this region. The ratio model-MAT comparison, in conjunction with the overall quality of the ratio model, indicates success in defining a powerful, yet simple enhanced pollen data model for employment in BHM contexts in central to northeastern temperate North America. Preliminary work (not shown) in coastal temperate northwestern North America with the ratio model western hemlock (*Tsuga heterophylla*) versus fir (*Abies*) suggests a model of similar quality can be developed for that region ($R^2 = 0.782$).

A ratio model of this kind could be emulated in a RSE context by defining two positive integer index values at selected climate model grid cells in a region of interest over a defined modern period, based on noisy negative and positive correlations (using absolute correlation values of ~ 0.3-0.5) with a seasonal or annual temperature mean. By scaling the sum of the index values to a number typical of pollen count data (a number between 100-300 would reasonably mimic typical pollen sums for the four taxa in the ratio model developed here), a paired set of simulated *n* and *r* values driven by temperature could thus be defined. A sequence of such paired *n* and *r* values could then be determined at multidecadal time steps, similar to the time resolution for simulated pollen data used by Li et al. (2010).

S.2 Use of the model in paleotemperature reconstruction and uncertainty estimation

The ratio model also has been tested for its capacity to directly reconstruct past temperature, outside a BHM context. This was done by solving the GLM equations for T as a function of α, β, and E(r|T). This solution is shown below, and reveals the logit "link function" in the GLM, of which the logistic EV function is the inverse.*

$$T = \{\ln[E(r|T)/\{1-E(r|T)\}]-\alpha\}/\beta$$

To enable use in paleoreconstruction, the observed *r* values from fossil pollen data are, by assumption, treated as unbiased estimators of the true *r* value at each time sampled, which therefore can be substituted for E(r|T). The formula then yields the expected values of T, given the observed r and estimated values for the parameters α and β. Uncertainty in this reconstruction can be evaluated in a systematic way in two parts: 1) the observed *r* values and their corresponding *n* values are input as the parameters into the binomial distribution, from which Monte Carlo random pollen count draws are made and a new ratio value determined for each iteration (again assuming unbiasedness in the observed ratio); and 2) the estimated standard errors for α and β are used with the α and β

* It is recognized that incorporation of the estimated forward model into a BHM context, e.g., represents a more sophisticated inference and is the ultimate goal of this work. Inversion of the model and use of the Monte Carlo framework outlined to estimate the EV and distribution of reconstructed T values were employed as a useful test of the capacity of the GLM model to provide meaningful temperature reconstructions. It was considered valuable to perform such a test prior to incorporation of the estimated model in a full BHM framework.

EV estimates to make a Monte Carlo set of random draws of α and β for each binomial draw (assuming α and β are Gaussian distributed). As an example, 1000 draws of α and β were made for each of 1000 binomial draws, generating one million uncertainty reconstructions of T to go along with the EV T reconstruction, for each of 39 fossil pollen samples covering the past 1000 years at a site in northwestern Wisconsin, USA (Gajewski et al. 1985; Gajewski 1988). The results are shown in Figure S2.

There are several features of interest in this paleotemperature reconstruction. From a technical standpoint, the nearly identical values for the EV reconstruction and the 50th percentile reconstruction from the Monte Carlo process indicate that the number of iterations is sufficient to ensure appropriate estimation of the conditional distribution of reconstructed T values for each fossil pollen sample. The EV reconstruction of the most recent sample (for 1968 21.1°) is very close to the actual 1960-1990 July temperature climatology at the site (21.2°), which satisfies a minimum requirement that the reconstruction calibration should reproduce the modern climatology within estimated uncertainties. The relative "narrowness" of the estimated uncertainty ranges arises from the high absolute ratios between the estimated values of α and β and their corresponding standard errors, which derives from the high explained deviance in the GLM model for the pollen types used. This reduction of uncertainty is an important measure of the value of the model selection process outlined. Slightly narrower uncertainty ranges might be derived if the ratio model is fitted on the entire data set, shown in Figure S3. Such a usage follows a convention of the dendroclimatological community: data partitions employed for model selection, calibration, and validation are then in some cases (limited a priori to situations where there is good quality in both calibration and validation)

relaxed for calibration of the model actually used in paleoclimate reconstruction, which employs the entire set of available data.

From a paleoclimatological standpoint, the reconstruction is interesting in that it shows important centennial-scale variability that is detectable as statistically significant due to the reduction of uncertainty (cf. Gajewski, 1988). The cooling and warming over the past 500 years is consistent with the well-known Little Ice Age (LIA) trajectory for the overall Northern Hemisphere mean temperature (cf. Mann et al. 2008), as is the particularly strong coolness reconstructed for the very-late 18th and early 19th centuries. [Note that timing of the LIA varies in paleoenvironmental reconstructions from different parts of the Northern Hemisphere. The onset of the LIA in this reconstruction is relatively late compared to the majority of last millennium reconstructions that are used to examine the timing of the entry into and exit out of the LIA, but is consistent with some of them.] A similar time trajectory of reconstructed temperatures also occurs for two nearby annually laminated lakes (Ruby Lake and Little Pine Lake, not shown; Gajewski et al. 1985; Gajewski 1988), indicating that this temperature pattern is likely not caused by unique, and possibly non-climatic, conditions at Dark Lake. Work is now in progress to extend these test reconstructions to further eastern sites in North America.

S.3 Technical note

The GLM algorithm in the R language was used for model fitting, reconstruction, and uncertainty estimation in this work. The power set analyses were done using Mathematica®. Mathematica® output was compared against test R output on identical data sets before starting the power set runs, which yielded identical results.

Figure S1 Results for the *Picea* plus *Betula* versus *Quercus* plus *Carya* ratio model. The top panel shows the spatial distribution of modern data sites used in GLM estimation, color-coded by the size of the residual between the observed and fitted ratio value at each site, normalized by the mean July temperature at the site. The bottom panels show, from left to right: a) the scatter plot of the ratio values versus temperature along with the GLM fitted E(r|T) values; b) the plot of fitted versus observed values, note the included 1/1 dotted line; and c) the plot of GLM Pearson residuals versus anomaly values of the predictor variable. The shading in the two left-hand bottom panels represents density of the data points derived from an empirical kernel estimator, shown as 25th, 50th, and 75th percentile values, from dark to light respectively. The red line in the right-hand bottom panel is a smoothed line fitted to the data points, using "panel.smooth" in R. The variability of the Pearson residuals is relatively, although not entirely, even in relation to the predictor variable anomalies, with one large outlier out of 465 data values.

Figure S2 Reconstruction of July mean temperature at Dark Lake, Wisconsin, USA based on inversion of the binomial logistic pollen model described in the text. Uncertainties are estimated from a two-way Monte Carlo process, taking random draws from the binomial distribution driven by the fossil pollen counts and from the Gaussian distribution driven by the estimated values of α and β and their standard errors, also described in the text. "Varve Years" are counts of (presumably) annual laminations in the sediment, going back from the time the sediments were cored at the site. Note the greater sampling density of pollen counts in the most recent 200 varve years. The fossil pollen data come from the Neotoma Paleoecology Database (http://www.neotomadb.org/).

Figure S3 Same as Figure S2, except that estimation of the binomial-logistic GLM model includes all modern data from the examination region. The R^2 of this fit is 0.793 (+ 0.011, - 0.100, ~95% level, from a bootstrap analysis that leaves out 10% of the data in each of 1000 iterations).

Figure S1

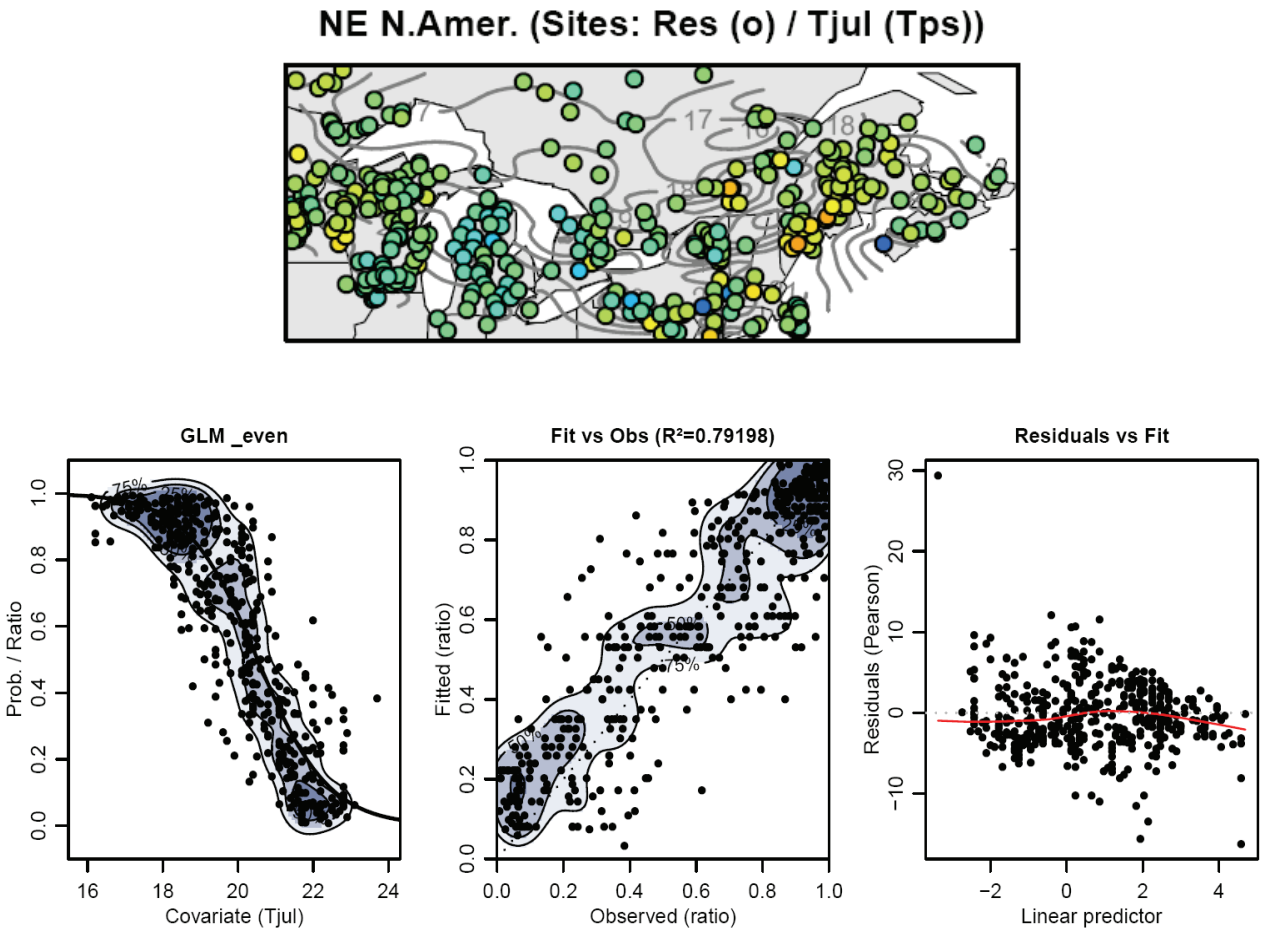

NE N.Amer. (Sites: Res (o) / Tjul (Tps))
GLM _even
Prob. / Ratio
Covariate (Tjul)
Fit vs Obs (R²=0.79198)
Fitted (ratio)
Observed (ratio)
Residuals vs Fit
Residuals (Pearson)
Linear predictor

Figure S2

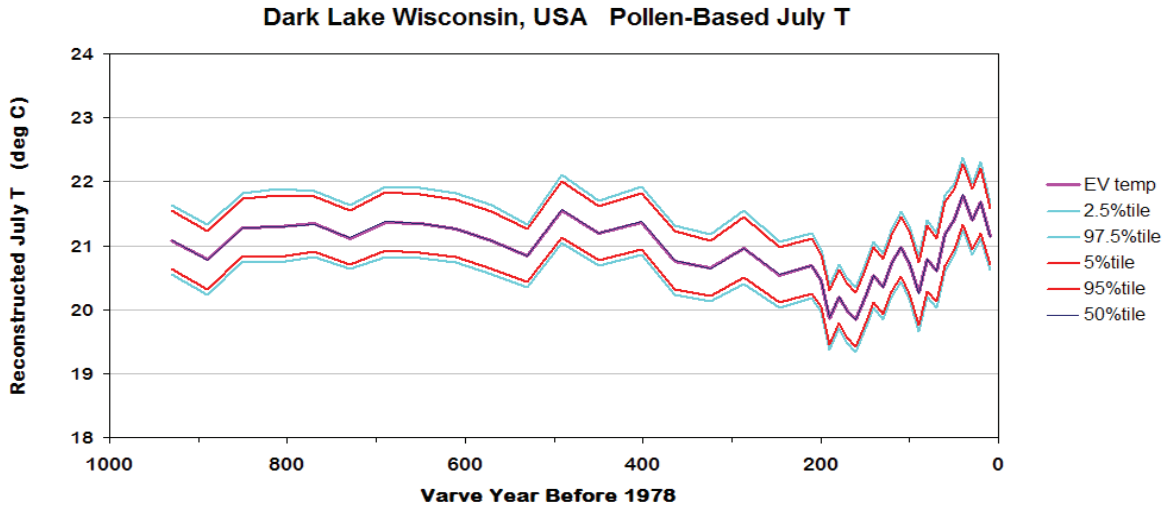
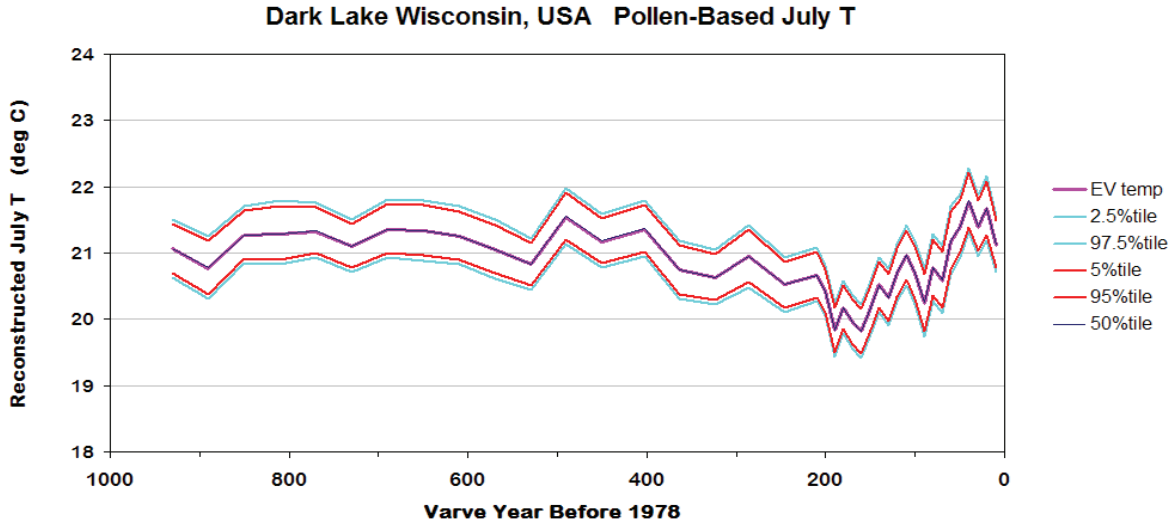


Figure S3



Supplemental Material References

Gavin, D., Oswald, W. W., Wahl, E. R., and Williams, J. (2003), "A Statistical Approach for Evaluating Distance Metrics and Analog Assignments for Pollen Records," *Quaternary Research*, 60:3, 356-367.

Gajewski, K., Winkler, M., and Swain, A. M.. (1985), "Vegetation and fire history from three lakes with varved sediments in northwestern Wisconsin (U.S.A.)", *Review of Paleobotany and Palynology*, 44, 277-292.

Gajewski, K. (1988), "Late-Holocene climates of eastern North America estimated from pollen data," *Quaternary Research* 29, 255-262.

Jackson, S. T., and Williams, J. W. (2004), "Modern Analogs in Quaternary Paleoecology: Here Today, Gone Yesterday, Gone Tomorrow?", *Annual Review of Earth and Planetary Science*, 32, 495-537, doi: 10.1146/annurev.earth.32.101802.120435.

Li, B., Nychka, D. W., and Ammann, C. M. (2010), "The Value of Multi-proxy Reconstruction of Past Climate," *Journal of the American Statistical Association*, in press, this issue.

Mann M. E., Zhang Z., Hughes M. K., Bradley R. S., Miller S. K., Rutherford S., and Ni, F. (2008), "Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia," *Proceedings of the National Academy of Science*, 105, 13252-13257, doi: 10.1073_pnas.0805721105.

Wahl, E. R. (2004), "A General Framework for Determining Cutoff Values to Select Pollen Analogs with Dissimilarity Metrics in the Modern Analog Technique," *Review of Palaeobotany and Palynology*, 128, 263-280.

Whitmore, J., Gajewski, K., Sawada, M., Williams, J.W., Shuman, B., Bartlein, P. J., Minckley, T., Viau, A. E., Webb, T. III, Shafer, S., Anderson, P., and Brubaker, L. (2005), "Modern Pollen Data from North American and Greenland for Multi-scale Paleoenvironmental Applications," *Quaternary Science Reviews*, 24, 1828-1848.

Williams, J. W., and Shuman, B. (2008), "Obtaining accurate and precise environmental reconstructions from the modern analog technique and North American surface pollen dataset," *Quaternary Science Reviews*, 27, 669-687.